● PRINTER RUSH ●
(PTO ASSISTANCE)

Application : 09/848,828   Examiner : BASHORE   GAU : 2176

From: M. DONATO   Location: (IDC) FMF FDC   Date: 01/03/06

Tracking #: EPM0984828   Week Date: 11/14/05

DOC CODE	DOC DATE	MISCELLANEOUS
☐ 1449	_____	☐ Continuing Data
☐ IDS	_____	☐ Foreign Priority
☐ CLM	_____	☐ Document Legibility
☐ IIFW	_____	☐ Fees
☐ SRFW	_____	☐ Other
☐ DRW	_____	
☐ OATH	_____	
☐ 312	_____	
☒ SPEC	05/04/01	

[RUSH] **MESSAGE:** _____

THERE IS MISSING INFORMATION ON PAGES
1, 5, 7, 10, 11 AND 13 OF THE SPECIFICATION.

THANK YOU.
M

[XRUSH] **RESPONSE:** _____

INITIALS:

NOTE: This form will be included as part of the official USPTO record, with the Response document coded as XRUSH.

REV 10/04

Patent   Docket No. RSW9-2001-0077-US1

DEDICATED PROCESSOR FOR EFFICIENT
PROCESSING OF DOCUMENTS ENCODED IN A MARKUP LANGUAGE

RELATED APPLICATIONS

The present invention is related to U.S. Patent No. 6938204, titled "Array-Based

Extensible Document Storage Format" (Application No. 09/652,296, and U.S. Patent No.

6941511, titled "High-Performance Extensible Document Transformation" (Application No.

09/653,080), and U.S. Patent No. 6904562, titled "Machine-Oriented Extensible

Document Representation And Interchange Notation" (Application No. 09/652,056), each

filed August 31, 2000. These related inventions are commonly assigned to International

Business Machines Corporation (IBM), and are hereby incorporated herein by reference.

FIELD OF THE INVENTION

The present invention relates generally to documents encoded in a markup language,

such as eXtensible Markup Language (XML), and particularly to processing of XML

documents in XML environments, such as a communications network.

In some embodiments, the document tree may be manipulated to create a document array model structure, as is generally known in the art. Generally, in an array model, data is organized to represent an ordered set of values that can be accessed by supplying one or more values which uniquely identify one of the values of the set. Accordingly, human-friendly

5   markup language tags are represented in an array model rather than a tree model. The array model simplifies and expedites processing.

In addition, XML documents can be transformed into or represented in the mXML language, a machine-oriented language similar to XML. U.S. Patent No. 6904562, titled "Machine-Oriented Extensible Document Representation And Interchange Notation" (Application No. 09/652,056), filed August 31, 2000, discloses the mXML notation. The mXML notation is more compact than the human-friendly XML notation and therefore provides performance gains in processing and transmission.

The parsing, transformation and other manipulation steps, e.g. XML document recognition, content based style sheet selection, content based routing and other traditional

15   XML processing steps, are tremendously processor intensive, which is burdensome on the general purpose processor and other system resources. Specifically, such processing steps prevent or delay the general purpose processor from performing other tasks required of the general purpose processor.

What is needed is a special purpose, dedicated processor for processing documents

20   encoded in a markup language such as XML which can free the general purpose processor to perform other tasks, and at least a hardware-based dedicated processor which can provide for optimization of processing steps by eliminating or reducing inefficiencies in human-friendly software code of the type heretofore known by relying on machine language characteristics.

example, one of several general purpose processors in a multi-processor computer system may be designated as the dedicated processor.

In either embodiment, the dedicated processor may be provided remotely, e.g. in a processing device which receives and processes documents before receipt by the intended

5   target. An arrangement is which the dedicated processor is network accessible has been found particularly advantageous because it is capable of supporting numerous devices and thereby offloading processing for numerous devices. Alternatively, in either a hardware- or software-based embodiment, the dedicated processor may be provided locally in the target device, e.g. co-located with a general purpose processor in a single device.

10   To achieve further performance benefits, the dedicated processor may optionally be configured to carry out XML processing using the array-based notation disclosed in U.S. Patent No. 693804, titled "Array-Based Extensible Document Storage Format" (Application No. 09/652,296, the transformation techniques disclosed in U.S. Patent No. 694151, titled "High-Performance Extensible Document Transformation" (Application No.

15   09/653,080), and the machine-oriented XML notation disclosed in U.S. Patent No. 690562, titled "Machine-Oriented Extensible Document Representation And Interchange Notation" (Application No. 09/652,056), each filed August 31, 2000.

The present invention provides a method for efficient processing of a document encoded in a markup language, the method comprising the step of communicating an array-

20   based data model representing the document to an application process through a bus of a printed circuit board. The present invention further provides a method for efficient processing of a document encoded in a markup language comprising the steps of receiving a document intended for delivery to a target, processing the document using a special purpose

hardware or software-based (as discussed further below), or whether the special purpose

processor is located locally or remotely, as discussed further below. This communication

also results regardless of whether the document is transformed or otherwise manipulated after

parsing, or a combination thereof.

5   The added overhead of the human-friendly tag syntax makes processing, e.g. parsing

to create the DOM tree, of the document burdensome to the general purpose processor. This

burden is unnecessary when the documents will only be "seen" by a computer program, such

as for those documents which are formatted for interchange between computer programs for

business-to-business ("B2B") or business-to-consumer ("B2C") use.

10   One way to improve processing efficiency is to abandon the human-friendly tag

structure. The assignee hereof has previously developed a machine-oriented notation for use

as an XML alternative. The machine-oriented notation improves processing time for

arbitrarily-structured documents and reduces the storage requirements and transmission costs

of data interchange while still retaining the extensibility and flexibility of XML and while

15   conveying equivalent content and semantic information. This machine-oriented notation is

referred to herein as "mXML". U.S. Patent No._6904562_, titled "Machine-Oriented

Extensible Document Representation And Interchange Notation" (Application No.

09/652,056), filed August 31, 2000 discloses the mXML notation, as well as a method,

system, and computer program product for operating upon (e.g. parsing, and storing

20   documents in) mXML. Accordingly, in a preferred embodiment, the dedicated processor is

configured to understand and interpret mXML, thereby resulting in processing efficiencies.

Creation of a DOM tree is computationally expensive in terms of processing time and

memory requirements. Using this tree-oriented DOM representation as an internal storage

format requires a considerable amount of memory and/or storage space to store the required

objects. In addition, a number of computer program instructions must be executed to allocate

memory and create the objects, delete objects and de-allocate memory, and traverse the tree

structure to perform operations thereon. Execution of these instructions increases the

5   processing time required for structured documents, as do the operating system-invoked

instructions which are periodically executed to perform "garbage collection" (whereby the

space being used by objects can be reclaimed after the objects have been logically deleted or

de-allocated).

Another way to improve processing efficiency is to use an array-based notation. The

10   Xalan XSLT (Extensible Language Transformations) processor from the Apache Software

Foundation reduces the number of objects used by DOM processors somewhat by providing

an in-memory Document Table Model ("DTM") representation of a DOM tree. An array is

used instead of a set of "real objects" for storing the DOM tree itself. However, there are still

many objects around to represent the XML data content of a document (including objects for

15   the nodes, node values, attributes, attribute values, etc.). Array-based processing makes it

easier to navigate the tree structure, e.g. for transformation purposes, etc. Accordingly, by

implementing array-based processing into the dedicated processor, further performance gains

are realized. In a highly preferred embodiment, the dedicated processor is configured to

process a document using the array-based notation disclosed in U.S. Patent No. 6938204 ,

20   titled "Array-Based Extensible Document Storage Format" (Application No. 09/652,296).

Figure 2A provides a flowchart 20 which sets forth a first embodiment of exemplary

logic for processing documents in accordance with Figure 1. In the example of Figure 2A, a

hardware-based special purpose processor is provided remotely, e.g. as a special purpose chip

software is capable of processing HTML, but not XML. Accordingly, a JAVA or other plug-in software application is typically executed by a general purpose processor within the device to translate the XML to HTML for post-processing, e.g. interpretation and display, by the web browser and general purpose processor. This places a burden on the general purpose

5   processor of the target devices to convert XML to HTML. Accordingly, in this example, server 346 is provided with a hardware-based special purpose processor for processing XML documents. In the example of Figure 2A, and as shown in Figure 3, an XML document deliverable to device 310a from data server 348 is first received (and implicitly recognized as such by a hardware or software based recognition engine) at an intermediate processing

10   device (server 346) as shown at step 22 of Figure 2A. The XML document is then processed, e.g. parsed by the hardware-based special processor of server 346, as shown at step 24 of Figure 2A. For example, such parsing results in creation of a document tree data model representing the XML document, e.g. in document object model (DOM) format. Alternatively, the special purpose processor of device 346 is configured parse the document

15   to create a data model in document array model (DAM) format. For example, a document array model may be created in accordance with the method described in U.S. Patent No. 6938204 , titled "Array-Based Extensible Document Storage Format" (Application No. 09/652,296).

        Optionally, e.g. if required for the target device, the document is further processed to

20   perform a transformation, as shown at step 26 of Figure 2A. For example, such transformations are typically performed to format content deliverable to handheld devices such as personal digital assistant (PDA) device 310b or web-enabled wireless telephone 310c of Figure 3. For example, such transformations are now typically performed by IBM's